Yinuo Ren
Joint work with Grant M. Rotskoff and Lexing Ying
Stanford University

March 24, 2026

A Unified Approach to Analysis and Design of Denoising Markov Models

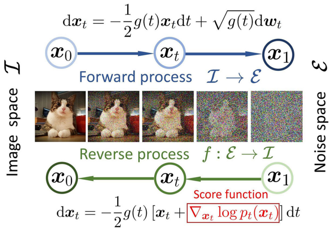
MSR New England Generative Modeling & Sampling Seminar

Outline

Section 1:
Introduction

Flow- and Diffusion-Based Generative Modeling

- **Diffusion models** learn reverse-time stochastic dynamics from a simple reference law back to data and now dominate many image, audio, and video generation tasks [SSDK+20].
- **Stochastic interpolants** and **flow matching** show that one can also learn along prescribed probability paths [ABVE23, LCBH+22, LGL22].



High-Level Recipe

One would like to design a *Markovian probability path* leading to the data distribution that

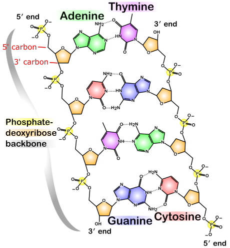
- originates from a simple reference law that are easy to sample from;
- has the explicit form to be simulated;
- can be learned from data via a tractable loss function.

Flow- and Diffusion-Based Generative Modeling

Variants

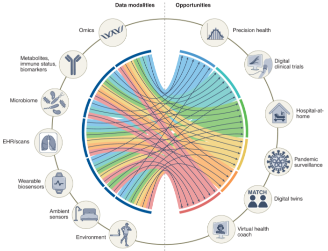
Continuous

- *Image, Video, Audio*
- (Continuous) Diffusion Models
- **Diffusion Process**



Discrete

- *Text, DNA sequence*
- Discrete Diffusion Models
- **Jump Process/CTMC**



Multi-Modal & Hard-Constrained

- *Txt2Img, Scientific Data*
- Specialized Models
- **General Markov Process**

(Continuous) Diffusion Models

- **Forward process:** Diffusion process on $\mathbb{R}^d$

$$\mathrm{d}\boldsymbol{x}_t = \boldsymbol{b}_t(\boldsymbol{x}_t)\,\mathrm{d}t + \mathrm{d}\boldsymbol{w}_t, \qquad \boldsymbol{x}_0 \sim p_0.$$

- **True backward process:** Also diffusion process on $\mathbb{R}^d$

$$\mathrm{d}\overleftarrow{\boldsymbol{x}}_t = \left(-\boldsymbol{b}_{T-t}(\overleftarrow{\boldsymbol{x}}_t) + \nabla \log p_{T-t}(\overleftarrow{\boldsymbol{x}}_t)\right)\mathrm{d}t + \mathrm{d}\boldsymbol{w}_t.$$

The density-dependent term is the score $\nabla \log p_t$ [And82].

- **Approximate backward process:** Approximate the score function $\nabla \log p_t$ by a neural network $\hat{\boldsymbol{s}}_t$ [SSDK+20]

$$\mathrm{d}\boldsymbol{y}_t = \left(-\boldsymbol{b}_{T-t}(\boldsymbol{y}_t) + \hat{\boldsymbol{s}}_{T-t}(\boldsymbol{y}_t)\right)\mathrm{d}t + \mathrm{d}\boldsymbol{w}_t, \qquad \hat{\boldsymbol{s}}_t \approx \nabla \log p_t.$$

- **Training principle:** Fit the approximate backward path law to the true backward path law in KL; this reduces to score matching [HD05]:

$$\mathbb{E}_{\boldsymbol{x}_0 \sim p_0}\left[\int_0^T \mathbb{E}_{\boldsymbol{x}_t \sim p_{t|0}(\cdot|\boldsymbol{x}_0)} \left\| \hat{\boldsymbol{s}}_t(\boldsymbol{x}_t) - \nabla \log p_{t|0}(\boldsymbol{x}_t \mid \boldsymbol{x}_0) \right\|^2 \mathrm{d}t\right].$$

Discrete Diffusion Models

- **Forward process:** Continuous-time Markov chain on a finite state space $\mathbb{X}$

$$\frac{\mathrm{d}\boldsymbol{p}_t}{\mathrm{d}t} = \boldsymbol{\Lambda}_t \boldsymbol{p}_t, \qquad \boldsymbol{\Lambda}_t = (\lambda_t(y,x))_{x,y\in\mathbb{X}},$$

  where $\lambda_t(y,x)$ is the jump rate from x to y [AJH+21, CBDB+22].

- **True backward process:** Also a continuous-time Markov chain on $\mathbb{X}$

$$\overleftarrow{\lambda}_{T-t}(y,x) = \frac{p_t(y)}{p_t(x)}\,\lambda_t(x,y).$$

  The density-dependent term is now a score ratio $\frac{p_t(y)}{p_t(x)}$ rather than a gradient.

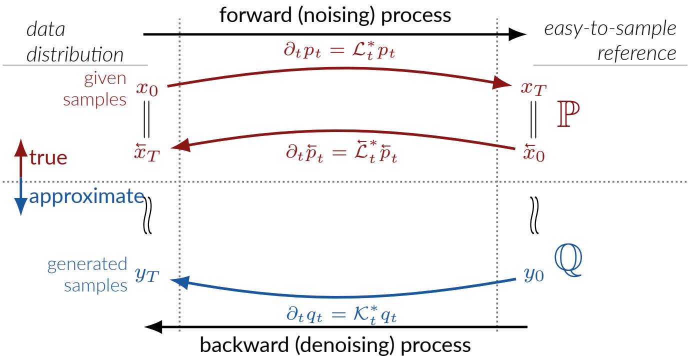
- **Approximate backward rates:** Approximate the score ratio $\frac{p_t(y)}{p_t(x)}$ by a neural network $\hat{s}_t$

$$\hat{\lambda}_{T-t}(y,x) = \hat{s}_t(x,y)\,\lambda_t(x,y), \qquad \hat{s}_t(x,y) \approx \frac{p_t(y)}{p_t(x)}.$$

- **Training principle:** Fit the approximate backward path law to the true backward path law in KL; this reduces to score matching [LME23, RCRY24, RCZ+25]:

$$\mathbb{E}\left[\int_0^T \sum_{y\neq x_t}\left(\hat{s}_t(x_t,y) - \frac{p_{t|0}(y\mid x_0)}{p_{t|0}(x_t\mid x_0)}\log \hat{s}_t(x_t,y)\right)\lambda_t(x_t,y)\ \mathrm{d}t\right].$$

Conceptual picture



True quantities: $(x_t, p_t, \overleftarrow{x}_t, \mathbb{P})$ are induced by the forward generator $\mathcal{L}_t$.
Approximate quantities: $(y_t, q_t, \mathcal{K}_t, \varphi_t, \mathbb{Q})$ define the approximate backward model.
Mismatch ratio: $\eta_t =: \varphi_t / p_t$ links the two and is the object behind both the Doob's h-transform and the training loss.

Observation

Ingredient	Continuous diffusion	Discrete diffusion
Forward dynamics	SDE / diffusion generator	CTMC / jump generator
True backward	depends on $\nabla \log p_t$	depends on $p_t(y)/p_t(x)$
Learned object	score field	score ratio field
Loss	L^2-loss	score-entropy loss

Question

- Is there a single framework that explains *continuous diffusion* and *discrete diffusion*?
- Can we design generative models with an arbitrary Markov process so as to enrich the design space and tailor to the properties of the data distribution?

Related perspectives

- [BSDB+24]: From denoising diffusion to denoising Markov models
- [HHY+24]: Generator matching
- **[RRY25]: A unified approach to analysis and design of denoising Markov models**

Section 2:
Mathematical Framework

Markov Processes and Generators

Definition (Generator)

For a Markov process $(x_t)_{t\in[0,T]}$ on a state space E, the time-dependent generator is

$$\mathcal{L}_t f(x) = \lim_{h\to 0^+} \frac{\mathbb{E}[f(x_{t+h}) \mid x_t = x] - f(x)}{h}, \quad \text{for} \quad f \in B_b(E).$$

- The forward marginals p_t evolve by the **Kolmogorov forward equation**

$$\partial_t p_t = \mathcal{L}_t^* p_t.$$

- For suitable test functions, **Dynkin's formula** gives

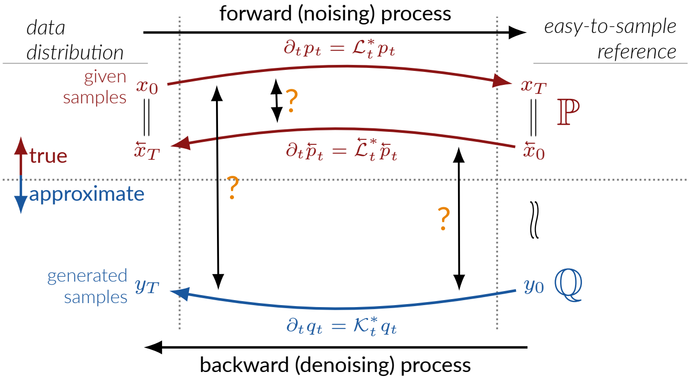
$$\mathbb{E}[f(x_t)] - f(x_0) = \mathbb{E}\left[\int_0^t \mathcal{L}_s f(x_s)\, \mathrm{d}s \middle| x_0\right].$$

- A key bilinear perturbation term is the carré-du-champ operator

$$\Gamma_t^*(u, f) =: \mathcal{L}_t^*(uf) - u\mathcal{L}_t^* f - f\mathcal{L}_t^* u,$$

which appears frequently in *stochastic analysis*, *functional inequalities*, and *nonequilibrium dynamics*.

Conceptual picture (Again)



Questions

- **Time-Reversal**: What is the relation between the forward generator $\mathcal{L}_t$ and its corresponding backward generator $\overleftarrow{\mathcal{L}}_t$?
- **Parameterization**: In which family should we search for the approximate backward generator $\mathcal{K}_t$?
- **Gap**: What is the relation between the approximate backward generator $\mathcal{K}_t$ and its truth $\overleftarrow{\mathcal{L}}_t$?

Generative Modeling with Markov Processes

Time-Reversal Process

Theorem (Time-reversal of the forward processes [CCGL23])

Under suitable regularity assumptions on the densities p_t, the generator of the time-reversal process $(\overleftarrow{x}_t)_{t\in[0,T]}$ satisfies

$$\overleftarrow{\mathcal{L}}_{T-t} f = \mathcal{L}_t^* f + p_t^{-1}\Gamma_t^*(p_t, f),$$

where $\Gamma_t^*(p_t, f) = \mathcal{L}_t^*(p_t f) - p_t \mathcal{L}_t^* f - f \mathcal{L}_t^* p_t$ is the carré-du-champ operator associated with the adjoint forward generator $\mathcal{L}_t^*$.

- The time-reversal generator $\overleftarrow{\mathcal{L}}_{T-t}$ is a perturbation to the adjoint generator $\mathcal{L}_t^*$ of the forward generator $\mathcal{L}_t$.
- The backward dynamics are explicit once we know the intermediate density p_t.

Generative Modeling with Markov Processes

Learning the Time-Reversal

Assumption (Parametrization of the backward generator)

We assume the approximate backward generator is parameterized by a strictly positive function $\varphi_t : E \to \mathbb{R}_+$ *through*

$$\mathcal{K}_{T-t} f = \mathcal{L}_t^* f + \varphi_t^{-1} \Gamma_t^* (\varphi_t, f),$$

where $\Gamma_t^*(\varphi_t, f) = \mathcal{L}_t^*(\varphi_t f) - \varphi_t \mathcal{L}_t^* f - f \mathcal{L}_t^* \varphi_t$ *is the carré-du-champ operator associated with the adjoint forward generator* $\mathcal{L}_t^*$.

- In practice, φ_t or an equivalent coordinate system is parameterized by a neural network.
- Instead of inventing a backward process from scratch, we restrict the search space via approximating the density-dependent perturbation in the exact backward generator.
- If $\varphi_t = p_t$, then $\mathcal{K}_t$ *exactly* equals the true backward generator.
- φ_t is the primitive quantity from which both continuous scores and discrete score ratios are derived.

Generative Modeling with Markov Processes

Quantifying the Gap to the True Time-Reversal

Lemma (Gap between the Estimated and True Backward Generators)

The approximate backward generator and the true backward generator satisfy

$$\mathcal{K}_{T-t} f = \overleftarrow{\mathcal{L}}_{T-t} f + \eta_t^{-1} \Gamma_t^*(\eta_t, f),$$

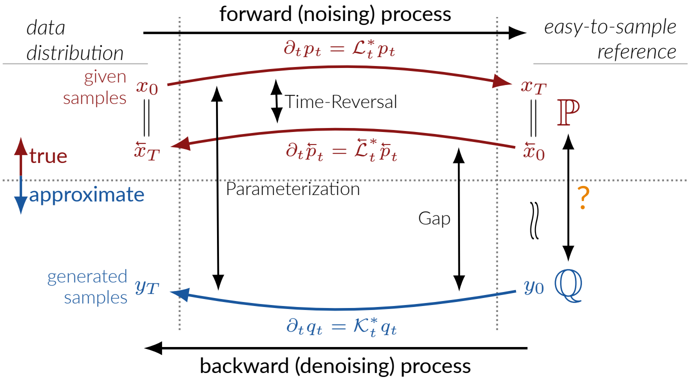
where we introduced the mismatch ratio

$$\eta_t =: \frac{\varphi_t}{p_t}$$

and $\Gamma_t^*(\eta_t, f) = \mathcal{L}_t^*(\eta_t f) - \eta_t \mathcal{L}_t^* f - f \mathcal{L}_t^* \eta_t$ *is the carré-du-champ operator associated with the adjoint forward generator* $\mathcal{L}_t^*$.

- The approximate backward generator $\mathcal{K}_{T-t}$ being a φ_t-perturbation to the adjoint forward generator $\mathcal{L}_t^* \Rightarrow \mathcal{K}_t$ being an η_t-perturbation to the true backward generator $\overleftarrow{\mathcal{L}}_t$.
- $\eta_t \equiv 1$ means perfect recovery of the true backward process.

Conceptual picture (Once Again)



- **Time-Reversal**: $\overleftarrow{\mathcal{L}}_{T-t} f = \mathcal{L}_t^* f + p_t^{-1} \Gamma_t^*(p_t, f)$.
- **Parameterization**: $\mathcal{K}_{T-t} f = \mathcal{L}_t^* f + \varphi_t^{-1} \Gamma_t^*(\varphi_t, f)$.
- **Gap**: $\mathcal{K}_{T-t} f = \overleftarrow{\mathcal{L}}_{T-t} f + \eta_t^{-1} \Gamma_t^*(\eta_t, f)$.
- **Question**: How to measure the gap between the true backward generator $\overleftarrow{\mathcal{L}}_{T-t}$ and the approximate backward generator $\mathcal{K}_{T-t}$ quantitatively?

Generalized Doob's h-transform

Theorem (Generalized Doob's h-transform [CT15])

Given a Markov generator $\mathcal{L}_t$, a positive function h_t, and a function λ_t, define

$$\mathcal{L}_t^{h,\lambda} f = h_t^{-1} \mathcal{L}_t(h_t f) - \lambda_t f.$$

Then the transformed process is absolutely continuous with respect to the original path measure, with Radon–Nikodym derivative

$$\frac{\mathrm{d}\mathbb{P}^{h,\lambda}}{\mathrm{d}\mathbb{P}}(x_{[0,T]}) = \frac{h_T(x_T)}{h_0(x_0)} \exp\left(- \int_0^T (\lambda_t + h_t^{-1} \partial_t h_t)(x_t) \, \mathrm{d}t \right),$$

where $x_{[0,T]}$ denotes the path of the process $(x_t)_{t \in [0,T]}$.

Our identification

- The estimated backward generator $\mathcal{K}_{T-t}$ is the generalized Doob's h-transform of the true backward generator $\overleftarrow{\mathcal{L}}_{T-t}$ with h_t equal to the mismatch ratio η_t.
- The choice $\lambda_t = \eta_t^{-1} \overleftarrow{\mathcal{L}}_{T-t} \eta_t$ makes the transformed generator conservative, and thus the approximate backward process is a path-space reweighting of the true backward process.

Change of Measure

By plugging in the time-reversal formula and Dynkin's formula for $\log \eta_t$, we can show the following change-of-measure theorem.

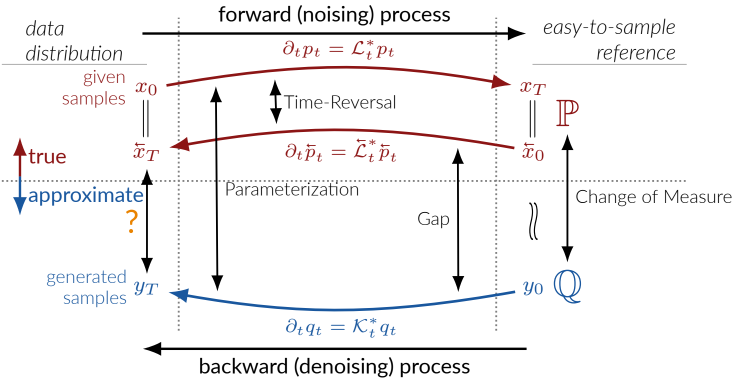
Theorem (Change of measure)

Under suitable regularity assumptions on η_t, there exists a probability measure $\mathbb{Q}$ such that the time-reversal process under $\mathbb{Q}$ is governed by $\mathcal{K}_t$, and

$$D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{Q}) = \mathbb{E}_{\mathbb{P}}\left[\int_0^T \left(\eta_t\, \mathcal{L}_t \eta_t^{-1} + \mathcal{L}_t \log \eta_t\right)(x_t)\ \mathrm{d}t\right] =: \mathfrak{L}[\eta_t].$$

- $\mathbb{P}$: path law of the forward process with generator $\mathcal{L}_t$ and the backward process with generator $\overleftarrow{\mathcal{L}}_t$.
- $\mathbb{Q}$: reweighted path law inducing the approximate backward generator $\mathcal{K}_t$.
- $\mathfrak{L}[\eta_t]$: the exact path-space discrepancy between the true and approximate backward dynamics.

Conceptual picture (Last Puzzle)



- **Change of Measure**: $D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{Q}) = \mathbb{E}\left[\int_0^T \left(\eta_t \mathcal{L}_t \eta_t^{-1} + \mathcal{L}_t \log \eta_t\right)(x_t)\,\mathrm{d}t\right]$.
- Question: How does this relate to the *training process* and *generation quality* of this generative model?

Error bound

Corollary (Error bound for the Approximate Backward Process)

The generated terminal law q_T satisfies

$$D_{\mathrm{KL}}(p_0 \| q_T) \le D_{\mathrm{KL}}(p_T \| q_0) + \mathfrak{L}[\eta_t].$$

- **Terminal mismatch** $D_{\mathrm{KL}}(p_T \| q_0)$: how well the forward process reaches the easy reference.
- **Estimation error** $\mathfrak{L}[\eta_t]$: how well the approximate backward process matches the true backward process.

Main message

Minimizing the training loss is not merely fitting a heuristic score. It directly controls an upper bound on the generative KL error.

Generalized Score-Matching

Define the conditional ratio

$$\eta_{t|0}(\cdot \mid x_0) =: \frac{\varphi_t(\cdot)}{p_{t|0}(\cdot \mid x_0)}.$$

Corollary (Generalized Score-Matching)

Up to constants independent of the model parameters,

$$\arg\min_\theta D_{\mathrm{KL}}(\mathbb{P}\|\mathbb{Q}^\theta) = \arg\min_\theta \mathfrak{L}[\eta_t^\theta] = \arg\min_\theta \mathfrak{L}_{\mathrm{SM}}[\eta_{t|0}^\theta],$$

where

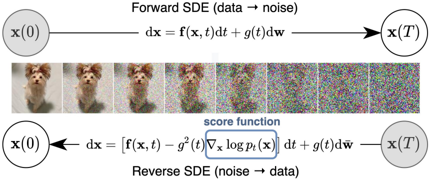
$$\mathfrak{L}_{\mathrm{SM}}[\eta_{t|0}] = \mathbb{E}_{x_0\sim p_0}\left[\int_0^T \mathbb{E}_{x_t\sim p_{t|0}(\cdot|x_0)}\Big(\eta_{t|0}\mathcal{L}_t\eta_{t|0}^{-1} + \mathcal{L}_t \log\eta_{t|0}\Big)(x_t)\ \mathrm{d}t\right].$$

- This enables learning the approximate backward generator $\mathcal{K}_t$ from data distribution p_0, forward transition kernel $p_{t|0}(\cdot|x_0)$, and the forward generator $\mathcal{L}_t$.

Meta-Algorithm

Needed ingredients

- Dataset $\hat{p}_0$
- Easy reference q_0
- Tractable conditional law $p_{t|0}$
- Simulator for $\mathcal{L}_t$



Training

1. Sample $x_0 \sim \hat{p}_0$ and time $t \sim \Psi$.
2. Sample $x_t \sim p_{t|0}(\cdot \mid x_0)$ by simulating the forward process.
3. Evaluate the integrand in $\mathfrak{L}_{\mathrm{SM}}$.
4. Update the parameters of φ_t^θ.

Inference

1. Sample $y_0 \sim q_0 \approx p_T$.
2. Simulate the approximate backward generator $\mathcal{K}_t^\theta$ from time 0 to T.
3. Output $y_T \sim q_T \approx p_0$.

Meta-Error Bound

Theorem (Meta-error bound for the Approximate Backward Process)

If the numerical simulation of the approximate backward process has one-step error $\mathcal{O}(\kappa^{1+r})$, then the generated law $\hat{q}_T$ satisfies

$$D_{\mathrm{KL}}(p_0 \| \hat{q}_T) \lesssim D_{\mathrm{KL}}(p_T \| q_0) + \mathfrak{L}(\eta_t^\theta) + T\kappa^r.$$

- **Truncation error**: $D_{\mathrm{KL}}(p_T \| q_0)$ from imperfect convergence of the forward process.
- **Estimation error**: $\mathfrak{L}(\eta_t^\theta)$ from learning the wrong backward dynamics.
- **Numerical error**: $T\kappa^r$ from discretizing the backward process.

Interpretation

The framework cleanly separates model design with numerical error.

Section 3:
Examples

Revisiting the Continuous Diffusion

- **State Space:** $E = \mathbb{R}^d$
- **Forward Generator:** Diffusion processes

$$\mathcal{L}_t f = \boldsymbol{b}_t \cdot \nabla f + \frac{1}{2} \boldsymbol{D}_t : \nabla^2 f$$

- **Backward Generator:**

$$\mathcal{K}_{T-t} f = \left(-\boldsymbol{b}_t + \boldsymbol{D}_t \nabla \log \varphi_t + \nabla \cdot \boldsymbol{D}_t \right) \cdot \nabla f + \frac{1}{2} \boldsymbol{D}_t : \nabla^2 f$$

- **Score Coordinates:** $\hat{\boldsymbol{s}}_t = \nabla \log \varphi_t$
- **Score-Matching Loss:**

$$\mathfrak{L}_{\mathrm{SM}}[\eta_{t|0}] = \mathbb{E}\left[\int_0^T \frac{1}{2} \boldsymbol{D}_t(\boldsymbol{x}_t) : \left(\hat{\boldsymbol{s}}_t(\boldsymbol{x}_t) - \nabla \log p_{t|0}(\boldsymbol{x}_t \mid \boldsymbol{x}_0) \right)^{\otimes 2} \mathrm{d}t \right]$$

Recovery of the classical case

When $\boldsymbol{D}_t = \sigma_t \boldsymbol{I}$, this is exactly the standard diffusion score-matching objective up to time reweighting [SSDK+20].

Revisiting the Discrete Diffusion

- **State Space:** $E = \mathbb{X}$
- **Forward Generator:** Jump processes on $\mathbb{X}$

$$\mathcal{L}_t f(x) = \sum_{y \in \mathbb{X}} (f(y) - f(x)) \lambda_t(y, x)$$

- **Backward Generator:**

$$\mathcal{K}_{T-t} f(x) = \sum_{y \in \mathbb{X}} (f(y) - f(x)) \hat{\lambda}_{T-t}(y, x), \quad \hat{\lambda}_{T-t}(y, x) = \frac{\varphi_t(y)}{\varphi_t(x)} \lambda_t(x, y)$$

- **Score Coordinates:** $\hat{s}_t(x, y) = \frac{\varphi_t(y)}{\varphi_t(x)}$
- **Score-Matching Loss:**

$$\mathfrak{L}_{\mathrm{SM}}[\eta_{t|0}] = \mathbb{E}\left[\int_0^T \sum_{y \neq x_t} \left(\hat{s}_t(x_t, y) - \frac{p_{t|0}(y \mid x_0)}{p_{t|0}(x_t \mid x_0)} \log \hat{s}_t(x_t, y)\right) \lambda_t(x_t, y) \, \mathrm{d}t\right]$$

Recovery of the classical case

This is precisely the discrete diffusion objective of [LME23, RCRY24], now seen as one coordinate system of the same generator-level framework.

Lévy-type models

Heavy-Tail Obstruction

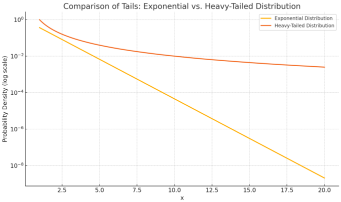
General KL guarantee: $D_{\mathrm{KL}}(p_0 \| q_T) \le D_{\mathrm{KL}}(p_T \| q_0) + \mathfrak{L}[\eta_t], \qquad \eta_t = \varphi_t / p_t.$

Theorem (Heavy-tail obstruction)

For affine Gaussian forward diffusions, if the data law p_0 is heavy-tailed with $\mathbb{E}\|\boldsymbol{x}_0\|^2 = \infty$ and the reference q_0 is Gaussian, then for every $T > 0$, we have

$$D_{\mathrm{KL}}(p_T \| q_0) = \infty.$$

- Conceptually, affine Gaussian forward dynamics do not repair a *tail mismatch* against a Gaussian terminal reference.
- This motivates forward processes with matching *tails and jump behavior*, *e.g.*, α-stable processes.



Lévy-type models

General Forward Generator

Theorem (Courrège form [Cou65, App09])

Under suitable regularity assumptions, the forward generator can be written as

$$\mathcal{L}_t f(\boldsymbol{x}) = \boldsymbol{b}_t(\boldsymbol{x}) \cdot \nabla f(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{D}_t(\boldsymbol{x}) : \nabla^2 f(\boldsymbol{x}) + \int_{\mathbb{R}^d \setminus \{\boldsymbol{x}\}} \Big(f(\boldsymbol{y}) - f(\boldsymbol{x}) - (\boldsymbol{y} - \boldsymbol{x}) \cdot \nabla f(\boldsymbol{x}) \, \mathbf{1}_{B(\boldsymbol{x},1)}(\boldsymbol{y} - \boldsymbol{x}) \Big) \lambda_t(\,\mathrm{d}\boldsymbol{y}, \boldsymbol{x}).$$

- $\boldsymbol{b}_t(\boldsymbol{x}) \cdot \nabla f$: deterministic transport of mass.
- $\frac{1}{2}\boldsymbol{D}_t(\boldsymbol{x}) : \nabla^2 f$: local Gaussian diffusion.
- The integral term: nonlocal jumps from $\boldsymbol{x}$ to $\boldsymbol{y}$; the linear correction on $B(\boldsymbol{x}, 1)$ compensates the infinitely many small jumps that are indistinguishable from the local transport.
- **Technical assumption**: the forward generator $\mathcal{L}_t$ acts on smooth compactly supported test functions [Cou65, App09].

Lévy-type models

Backward Triplet and Loss Decomposition

Estimated backward Lévy triplet

$$\overleftarrow{\lambda}^{\varphi}_{T-t}(\boldsymbol{y},\boldsymbol{x}) = \frac{\varphi_t(\boldsymbol{y})}{\varphi_t(\boldsymbol{x})}\,\lambda_t(\boldsymbol{x},\boldsymbol{y}),$$

$$\overleftarrow{\boldsymbol{D}}^{\varphi}_{T-t}(\boldsymbol{x}) = \boldsymbol{D}_t(\boldsymbol{x}),$$

$$\overleftarrow{\boldsymbol{b}}^{\varphi}_{T-t}(\boldsymbol{x}) = -\boldsymbol{b}_t(\boldsymbol{x}) + \boldsymbol{D}_t(\boldsymbol{x})\nabla\log\varphi_t(\boldsymbol{x}) + \nabla\cdot\boldsymbol{D}_t(\boldsymbol{x}) + \int_{B(\boldsymbol{x},1)}\left(\frac{\varphi_t(\boldsymbol{y})}{\varphi_t(\boldsymbol{x})}\lambda_t(\boldsymbol{x},\boldsymbol{y}) + \lambda_t(\boldsymbol{y},\boldsymbol{x})\right)(\boldsymbol{y}-\boldsymbol{x})\,\mathrm{d}\boldsymbol{y}.$$

- Reverse jumps are reweighted by the ratio $\varphi_t(\boldsymbol{y})/\varphi_t(\boldsymbol{x})$.
- The extra drift is the small-jump compensation term.
- *cf.* [YPKL23].

Path-space KL loss

$$\mathfrak{L}[\eta_t] = \underbrace{\mathbb{E}\left[\int_0^T \frac{1}{2}\,\boldsymbol{D}_t(\boldsymbol{x}_t) : \nabla\log\eta_t(\boldsymbol{x}_t)\nabla^\top\log\eta_t(\boldsymbol{x}_t)\,\mathrm{d}t\right]}_{\text{diffusion term}} + \underbrace{\mathbb{E}\left[\int_0^T\int_{\mathbb{R}^d}\ell\left(\frac{\eta_t(\boldsymbol{x}_t)}{\eta_t(\boldsymbol{y})}\right)\lambda_t(\boldsymbol{y},\boldsymbol{x}_t)\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}t\right]}_{\text{jump term}}, \qquad \ell(r) = r - 1 - \log r.$$

Experiment 1: Geometric Brownian Motion

Setup

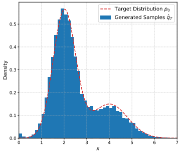
- Forward process:

$$\mathrm{d}\boldsymbol{x}_t = \boldsymbol{x}_t \odot \boldsymbol{\Sigma}\, \mathrm{d}\boldsymbol{w}_t, \quad \text{with} \quad \boldsymbol{D}(\boldsymbol{x}_t) = \operatorname{diag}(\boldsymbol{x}_t)\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top \operatorname{diag}(\boldsymbol{x}_t).$$

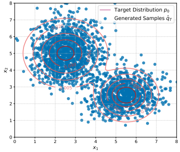
- Approximate backward:

$$\mathrm{d}\boldsymbol{y}_t = \big(\boldsymbol{D}(\boldsymbol{y}_t)\hat{\boldsymbol{s}}_{T-t}(\boldsymbol{y}_t) + \nabla \cdot \boldsymbol{D}(\boldsymbol{y}_t)\big)\mathrm{d}t + \boldsymbol{y}_t \odot \boldsymbol{\Sigma}\mathrm{d}\boldsymbol{w}_t.$$

- State space is the positive orthant $\mathbb{R}_+^d$: positivity is preserved by construction.
- Target data are mixtures of Gaussians restricted to $\mathbb{R}_+^d$.
- Approximate with the generalized diffusion score-matching loss.



(a) 1D result



(b) 2D result

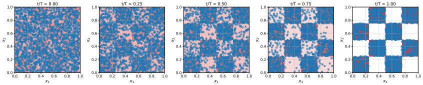
Observed behavior

Training remains stable even though the score can be singular near the boundary, and generated samples match the target without boundary violations. See also recent work [GRC25].

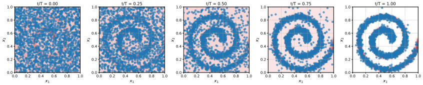
Experiment 2: Pure Jump Model

Setup

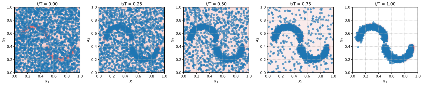
- Forward process: pure jump model with $\boldsymbol{b}_t \equiv 0$ and $\boldsymbol{D}_t \equiv 0$.
- Domain: the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$.
- Forward Lévy kernel: isotropic Gaussian jump kernel.
- Initial backward law q_0: approximately uniform on the torus.



(a) Chessboard



(b) Swiss roll



(c) Moons

Observed behavior

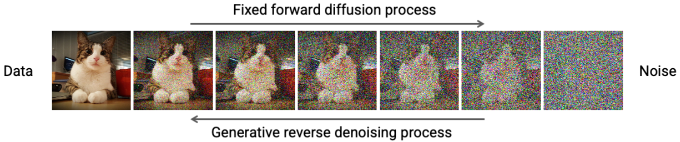
Samples start from a uniform law and progressively *jump* into structured target distributions, showing that the framework can learn meaningful backward dynamics without any diffusion component. See also a related example in [HHY$^+$24].

Section 4:
Conclusion

Recap

To build a generative model from an arbitrary Markov process, we need to:

1. Choose a forward Markov process with generator $\mathcal{L}_t$ that sends p_0 toward an easy reference q_0.
2. The exact backward generator is determined by the time-reversal theorem and depends on the current density p_t.
3. Parameterize the approximate backward generator by a positive function φ_t in the same structural form as the true backward generator.
4. The mismatch ratio $\eta_t = \varphi_t / p_t$ induces a change-of-measure theorem connecting the Doob's h-transform and the path-space KL objective.



Discussion

What Next?

- Better understanding of how forward dynamics affect *trainability*, *efficiency*, and *robustness*;
- Scalable inference and numerical methods adapted to general Markov/Lévy-type processes;
- Principled selection of forward processes tailored to constraints, topology, and distributional characteristics;
- Further enrich the design space of generative models beyond the current scope.

References I

[ABVE23] Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden, *Stochastic interpolants: A unifying framework for flows and diffusions*, arXiv preprint arXiv:2303.08797 (2023).

[AJH+21] Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg, *Structured denoising diffusion models in discrete state-spaces*, Advances in Neural Information Processing Systems **34** (2021), 17981–17993.

[And82] Brian DO Anderson, *Reverse-time diffusion equation models*, Stochastic Processes and their Applications **12** (1982), no. 3, 313–326.

[App09] David Applebaum, *Lévy processes and stochastic calculus*, Cambridge university press, 2009.

[BSDB+24] Joe Benton, Yuyang Shi, Valentin De Bortoli, George Deligiannidis, and Arnaud Doucet, *From denoising diffusions to denoising markov models*, Journal of the Royal Statistical Society Series B: Statistical Methodology **86** (2024), no. 2, 286–301.

[CBDB+22] Andrew Campbell, Joe Benton, Valentin De Bortoli, Thomas Rainforth, George Deligiannidis, and Arnaud Doucet, *A continuous time framework for discrete denoising models*, Advances in Neural Information Processing Systems **35** (2022), 28266–28279.

[CCGL23] Patrick Cattiaux, Giovanni Conforti, Ivan Gentil, and Christian Léonard, *Time reversal of diffusion processes under a finite entropy condition*, Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques, vol. 59, Institut Henri Poincaré, 2023, pp. 1844–1881.

References II

[Cou65] Philippe Courrège, *Sur la forme intégro-différentielle des opérateurs de* c_k^∞ *dans* c *satisfaisant au principe du maximum*, Séminaire Brelot-Choquet-Deny. Théorie du Potentiel **10** (1965), no. 1, 1–38.

[CT15] Raphaël Chetrite and Hugo Touchette, *Nonequilibrium markov processes conditioned on large deviations*, Annales Henri Poincaré, vol. 16, Springer, 2015, pp. 2005–2057.

[GRC25] Robert Gruhlke, Valentin Resseguier, and TALLA MAKOUGNE Merveille Cyndie, *Multiplicative diffusion models: Beyond gaussian latents*, The Fourteenth International Conference on Learning Representations, 2025.

[HD05] Aapo Hyvärinen and Peter Dayan, *Estimation of non-normalized statistical models by score matching.*, Journal of Machine Learning Research **6** (2005), no. 4.

[HHY+24] Peter Holderrieth, Marton Havasi, Jason Yim, Neta Shaul, Itai Gat, Tommi Jaakkola, Brian Karrer, Ricky TQ Chen, and Yaron Lipman, *Generator matching: Generative modeling with arbitrary markov processes*, arXiv preprint arXiv:2410.20587 (2024).

[LCBH+22] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le, *Flow matching for generative modeling*, arXiv preprint arXiv:2210.02747 (2022).

[LGL22] Xingchao Liu, Chengyue Gong, and Qiang Liu, *Flow straight and fast: Learning to generate and transfer data with rectified flow*, arXiv preprint arXiv:2209.03003 (2022).

[LME23] Aaron Lou, Chenlin Meng, and Stefano Ermon, *Discrete diffusion modeling by estimating the ratios of the data distribution*, arXiv preprint arXiv:2310.16834 (2023).

References III

[RCRY24] Yinuo Ren, Haoxuan Chen, Grant M Rotskoff, and Lexing Ying, *How discrete and continuous diffusion meet: Comprehensive analysis of discrete diffusion models via a stochastic integral framework*, arXiv preprint arXiv:2410.03601 (2024).

[RCZ+25] Yinuo Ren, Haoxuan Chen, Yuchen Zhu, Wei Guo, Yongxin Chen, Grant M Rotskoff, Molei Tao, and Lexing Ying, *Fast solvers for discrete diffusion models: Theory and applications of high-order algorithms*, arXiv preprint arXiv:2502.00234 (2025).

[RRY25] Yinuo Ren, Grant M Rotskoff, and Lexing Ying, *A unified approach to analysis and design of denoising markov models*, arXiv preprint arXiv:2504.01938 (2025).

[SSDK+20] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, *Score-based generative modeling through stochastic differential equations*, arXiv preprint arXiv:2011.13456 (2020).

[YPKL23] Eun Bi Yoon, Keehun Park, Sungwoong Kim, and Sungbin Lim, *Score-based generative models with lévy processes*, Advances in Neural Information Processing Systems **36** (2023), 40694–40707.

Thank you!

Questions?